

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Prudent, James R.  
Hall, Jeff G.  
Lyamichev, Victor I.
- (ii) TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
- (iii) NUMBER OF SEQUENCES: 69
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Medlen & Carroll, LLP
  - (B) STREET: 220 Montgomery Street, Suite 2200
  - (C) CITY: San Francisco
  - (D) STATE: California
  - (E) COUNTRY: United States Of America
  - (F) ZIP: 94104
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/682,853
  - (B) FILING DATE: 12-JUL-1996
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/599,491
  - (B) FILING DATE: 24-JAN-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Ingolia, Diane E.
  - (B) REGISTRATION NUMBER: 40,027
  - (C) REFERENCE/DOCKET NUMBER: FORS-02564
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (415) 705-8410
  - (B) TELEFAX: (415) 397-8338

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2506 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGAGGGGGA TGCTGCCCCT CTTTGAGCCC AAGGGCCGGG TCCTCCTGGT GGACGGCCAC	60
CACCTGGCCT ACCGCACCTT CCACGCCCTG AAGGGCCTCA CCACCAGCCG GGGGGAGCCG	120
GTGCAGGCGG TCTACGGCTT CGCCAAGAGC CTCCTCAAGG CCCTCAAGGA GGACGGGGAC	180
GCGGTGATCG TGGTCTTTGA CGCCAAGGCC CCTTCCTTCC GCCACGAGGC CTACGGGGGG	240
TACAAGGCGG GCCGGGCCCC CACGCCGGAG GACTTTCCCC GGCAACTCGC CCTCATCAAG	300
GAGCTGGTGG ACCTCCTGGG GCTGGCGCGC CTCGAGGTCC CGGGCTACGA GGCGGACGAC	360
GTCCTGGCCA GCCTGGCCAA GAAGGCGGAA AAGGAGGGCT ACGAGGTCCG CATCCTCACC	420
GCCGACAAAG ACCTTTACCA GTCCTTTTCC GACCGCATCC ACGTCCTCCA CCCCAGGGGG	480
TACCTCATCA CCCCAGCCTG GCTTTGGGAA AAGTACGGCC TGAGGCCCCGA CCAGTGGGCC	540
GACTACCGGG CCCTGACCGG GGACGAGTCC GACAACCTTC CCGGGGTCAA GGGCATCGGG	600
GAGAAGACGG CGAGGAAGCT TCTGGAGGAG TGGGGGAGCC TGGAAGCCCT CCTCAAGAAC	660
CTGGACCGGC TGAAGCCCGC CATCCGGGAG AAGATCCTGG CCCACATGGA CGATCTGAAG	720
CTCTCCTGGG ACCTGGCCAA GGTGCGCACC GACCTGCCCC TGGAGGTGGA CTTCGCCAAA	780
AGGCGGGAGC CCGACCGGGA GAGGCTTAGG GCCTTTCTGG AGAGGCTTGA GTTTGGCAGC	840
CTCTCCACG AGTTCGGCCT TCTGGAAAGC CCAAGGCCC TGGAGGAGGC CCCCTGGCCC	900
CCGCCGAAG GGGCCTTCGT GGGCTTTGTG CTTTCCCGCA AGGAGCCCAT GTGGGCCGAT	960
CTTCTGGCCC TGGCCGCCGC CAGGGGGGGC CGGGTCCACC GGGCCCCCGA GCCTTATAAA	1020
GCCCTCAGGG ACCTGAAGGA GGC GCGGGGG CTTCTCGCCA AAGACCTGAG CGTTCTGGCC	1080
CTGAGGGAAG GCCTTGGCCT CCCGCCCGGC GACGACCCCA TGCTCCTCGC CTACCTCCTG	1140
GACCCTTCCA ACACCACCCC CGAGGGGGTG GCGGCGGCT ACGGCGGGGA GTGGACGGAG	1200
GAGGCGGGGG AGCGGGCCGC CCTTCCGAG AGGCTCTTCG CCAACCTGTG GGGGAGGCTT	1260
GAGGGGGAGG AGAGGCTCCT TTGGCTTTAC CGGGAGGTGG AGAGGCCCTT TTCCGCTGTC	1320
CTGGCCACA TGGAGGCCAC GGGGGTGC GC CTGGACGTGG CCTATCTCAG GGCCTTGTCC	1380
CTGGAGGTGG CCGAGGAGAT CGCCCGCCTC GAGGCCGAGG TCTTCCGCCT GGCCGGCCAC	1440
CCCTTCAACC TCAACTCCCC GGACCAGCTG GAAAGGGTCC TCTTTGACGA GCTAGGGCTT	1500
CCCGCCATCG GCAAGACGGA GAAGACCGGC AAGCGCTCCA CCAGCGCCGC CGTCCTGGAG	1560
GCCCTCCGCG AGGCCACCC CATCGTGGAG AAGATCCTGC AGTACCGGGA GCTACCAAG	1620
CTGAAGAGCA CCTACATTGA CCCCTTGCCG GACCTCATCC ACCCCAGGAC GGGCCGCCCTC	1680
CACACCCGCT TCAACCAGAC GGCCACGGCC ACGGGCAGGC TAAGTAGCTC CGATCCCAAC	1740
CTCCAGAACA TCCCCGTCCG CACCCCGCTT GGGCAGAGGA TCCGCCGGGC CTTCATCGCC	1800

GAGGAGGGGT GGCTATTGGT GGCCCTGGAC TATAGCCAGA TAGAGCTCAG GGTGCTGGCC 1860  
CACCTCTCCG GCGACGAGAA CCTGATCCGG GTCTTCCAGG AGGGGCGGGA CATCCACACG 1920  
GAGACCGCCA GCTGGATGTT CGGCGTCCCC CGGGAGGCCG TGGACCCCTT GATGCGCCGG 1980  
GCGGCCAAGA CCATCAACTT CGGGGTCCCTC TACGGCATGT CGGCCCACCG CCTCTCCCAG 2040  
GAGCTAGCCA TCCCTTACGA GGAGGCCAG GCCTTCATTG AGCGCTACTT TCAGAGCTTC 2100  
CCCAAGGTGC GGGCCTGGAT TGAGAAGACC CTGGAGGAGG GCAGGAGGCG GGGGTACGTG 2160  
GAGACCCTCT TCGGCCGCCG CCGCTACGTG CCAGACCTAG AGGCCCGGGT GAAGAGCGTG 2220  
CGGGAGGCCG CCGAGCGCAT GGCCTTCAAC ATGCCCGTCC AGGGCACCGC CGCCGACCTC 2280  
ATGAAGCTGG CTATGGTGAA GCTCTTCCCC AGGCTGGAGG AAATGGGGGC CAGGATGCTC 2340  
CTTCAGGTCC ACGACGAGCT GGTCTTCGAG GCCCCAAAAG AGAGGGCGGA GGCCGTGGCC 2400  
CGGCTGGCCA AGGAGGTCAT GGAGGGGGTG TATCCCCTGG CCGTGCCCCCT GGAGGTGGAG 2460  
GTGGGGATAG GGGAGGACTG GCTCTCCGCC AAGGAGTGAT ACCACC 2506

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2496 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGGCGATGC TTCCCCTCTT TGAGCCCAA GGCCGCGTGC TCCTGGTGGA CGGCCACCAC 60  
CTGGCCTACC GCACCTTCTT TGCCCTCAAG GGCCTCACCA CCAGCCGCGG CGAACCCGTT 120  
CAGGCGGTCT ACGGCTTCGC CAAAAGCCTC CTCAAGGCC TGAAGGAGGA CGGGGACGTG 180  
GTGGTGGTGG TCTTTGACGC CAAGGCCCCC TCCTTCCGCC ACGAGGCCTA CGAGGCCTAC 240  
AAGGCGGGCC GGGCCCCCAC CCCGGAGGAC TTTCCCCGGC AGCTGGCCCT CATCAAGGAG 300  
TTGGTGGACC TCCTAGGCCT TGTGCGGCTG GAGGTTCCCG GCTTTGAGGC GGACGACGTG 360  
CTGGCCACCC TGGCCAAGCG GCGGAAAAG GAGGGGTACG AGGTGCGCAT CCTACTGCC 420  
GACCGCGACC TCTACCAGCT CCTTTCGGAG CGCATCGCCA TCCTCCACCC TGAGGGGTAC 480  
CTGATCACCC CGGCGTGGCT TTACGAGAAG TACGGCCTGC GCGCGGAGCA GTGGGTGGAC 540  
TACCGGGCCC TGGCGGGGGA CCCCTCGGAT AACATCCCCG GGGTGAAGGG CATCGGGGAG 600  
AAGACCGCCC AGAGGCTCAT CCGCGAGTGG GGGAGCCTGG AAAACCTCTT CCAGCACCTG 660  
GACCAGGTGA AGCCCTCCTT GCGGGAGAAG CTCCAGGCGG GCATGGAGGC CCTGGCCCTT 720  
TCCCGGAAGC TTTCCAGGT GCACACTGAC CTGCCCCTGG AGGTGGACTT CGGGAGGCGC 780

2006 年 12 月 22 日

CGCACACCCA	ACCTGGAGGG	TCTGCGGGCT	TTTTTGGAGC	GGTTGGAGTT	TGGAAGCCTC	840
CTCCACGAGT	TCGGCCTCCT	GGAGGGGCGG	AAGGCGGCAG	AGGAGGCCCC	CTGGCCCCCT	900
CCGGAAGGGG	CTTTTTTGGG	CTTTTCCTTT	TCCCGTCCCG	AGCCCATGTG	GGCCGAGCTT	960
CTGGCCCTGG	CTGGGGCGTG	GGAGGGGCGC	CTCCATCGGG	CACAAGACCC	CCTTAGGGGC	1020
CTGAGGGACC	TTAAGGGGGT	GCGGGGAATC	CTGGCCAAGG	ACCTGGCGGT	TTTGGCCCTG	1080
CGGGAGGGCC	TGGACCTCTT	CCCAGAGGAC	GACCCCATGC	TCCTGGCCTA	CCTTCTGGAC	1140
CCCTCCAACA	CCACCCCTGA	GGGGGTGGCC	CGGCGTTACG	GGGGGAGTG	GACGGAGGAT	1200
GCGGGGGAGA	GGGCCCTCCT	GGCCGAGCGC	CTCTTCCAGA	CCCTAAAGGA	GCGCCTTAAG	1260
GGAGAAGAAC	GCCTGCTTTG	GCTTTACGAG	GAGGTGGAGA	AGCCGCTTTC	CCGGGTGTTG	1320
GCCCGGATGG	AGGCCACGGG	GGTCCGGCTG	GACGTGGCCT	ACCTCCAGGC	CCTCTCCCTG	1380
GAGGTGGAGG	CGGAGGTGCG	CCAGCTGGAG	GAGGAGGTCT	TCCGCCTGGC	CGGCCACCCC	1440
TTCAACCTCA	ACTCCCGCGA	CCAGCTGGAG	CGGGTGCTCT	TTGACGAGCT	GGGCCTGCCT	1500
GCCATCGGCA	AGACGGAGAA	GACGGGGAAA	CGCTCCACCA	GCGCTGCCGT	GCTGGAGGCC	1560
CTGCGAGAGG	CCCACCCCAT	CGTGGACCGC	ATCCTGCAGT	ACCGGGAGCT	CACCAAGCTC	1620
AAGAACACCT	ACATAGACCC	CCTGCCCCGCC	CTGGTCCACC	CCAAGACCGG	CCGGCTCCAC	1680
ACCCGC'TTCA	ACCAGACGGC	CACCGCCACG	GGCAGGCTTT	CCAGCTCCGA	CCCCAACCTG	1740
CAGAACATCC	CCGTGCGCAC	CCCTCTGGGC	CAGCGCATCC	GCCGAGCCTT	CGTGGCCGAG	1800
GAGGGCTGGG	TGCTGGTGGT	CTTGGACTAC	AGCCAGATTG	AGCTTCGGGT	CCTGGCCCAC	1860
CTCTCCGGGG	ACGAGAACCT	GATCCGGGTC	TTTCAGGAGG	GGAGGGACAT	CCACACCCAG	1920
ACCGCCAGCT	GGATGTTTCG	CGTTTCCCCC	GAAGGGGTAG	ACCTCTTGAT	GCGCCGGGCG	1980
GCCAAGACCA	TCAACTTCGG	GGTGCTCTAC	GGCATGTCCG	CCCACCGCCT	CTCCGGGGAG	2040
CTTTCCATCC	CCTACGAGGA	GGCGGTGGCC	TTCATTGAGC	GCTACTTCCA	GAGCTACCCC	2100
AAGGTGCGGG	CCTGGATTGA	GGGGACCCTC	GAGGAGGGCC	GCCGGCGGGG	GTATGTGGAG	2160
ACCCTCTTCG	GCCGCCGGCG	CTATGTGCCC	GACCTCAACG	CCCGGGTGAA	GAGCGTGCGC	2220
GAGGCGGCGG	AGCGCATGGC	CTTCAACATG	CCGGTCCAGG	GCACCGCCGC	CGACCTCATG	2280
AAGCTGGCCA	TGGTGCGGCT	TTTCCCCCGG	CTTCAGGAAC	TGGGGGCGAG	GATGCTTTTG	2340
CAGGTGCACG	ACGAGCTGGT	CCTCGAGGCC	CCCAAGGACC	GGGCGGAGAG	GGTAGCCGCT	2400
TTGGCCAAGG	AGGTCATGGA	GGGGGTCTGG	CCCCTGCAGG	TGCCCCTGGA	GGTGGAGGTG	2460
GGCCTGGGGG	AGGACTGGCT	CTCCGCCAAG	GAGTAG			2496

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2504 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGAGGCGA TGCTTCCGCT CTTTGAACCC AAAGGCCGGG TCCTCCTGGT GGACGGCCAC	60
CACCTGGCCT ACCGCACCTT CTTGCCCCTG AAGGGCCTCA CCACGAGCCG GGGCGAACCG	120
GTGCAGGCGG TCTACGGCTT CGCCAAGAGC CTCTCAAGG CCCTGAAGGA GGACGGGTAC	180
AAGGCCGTCT TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGAG	240
GCCTACAAGG CGGGGAGGGC CCCGACCCCC GAGGACTTCC CCCGGCAGCT CGCCCTCATC	300
AAGGAGCTGG TGGACCTCCT GGGGTTTACC CGCCTCGAGG TCCCCGGCTA CGAGGCGGAC	360
GACGTTCTCG CCACCCTGGC CAAGAAGGCG GAAAAGGAGG GGTACGAGGT GCGCATCCTC	420
ACCGCCGACC GCGACCTCTA CCAACTCGTC TCCGACCGCG TCGCCGTCCT CCACCCCGAG	480
GGCCACCTCA TCACCCCGGA GTGGCTTTGG GAGAAGTACG GCCTCAGGCC GGAGCAGTGG	540
GTGGACTTCC GCGCCCTCGT GGGGGACCCC TCCGACAACC TCCCCGGGGT CAAGGGCATC	600
GGGGAGAAGA CCGCCCTCAA GCTCCTCAAG GAGTGGGGAA GCCTGGAAAA CCTCCTCAAG	660
AACCTGGACC GGGTAAAGCC AGAAAACGTC CGGGAGAAGA TCAAGGCCCA CCTGGAAGAC	720
CTCAGGCTCT CTTGGAGCT CTCCCGGGTG CGCACCGACC TCCCCCTGGA GGTGGACCTC	780
GGCCAGGGGC GGGAGCCCGA CCGGGAGGGG CTTAGGGCCT TCCTGGAGAG GCTGGAGTTC	840
GGCAGCCTCC TCCACGAGTT CGGCCTCCTG GAGGCCCCCG CCCCCCTGGA GGAGGCCCCC	900
TGGCCCCCGC CGGAAGGGGC CTTCTGTTGG TTCGTCCTCT CCCGCCCCGA GCCCATGTGG	960
GCGGAGCTTA AAGCCCTGGC CGCCTGCAGG GACGGCCGGG TGCACCGGGC AGCAGACCCC	1020
TTGGCGGGGC TAAAGGACCT CAAGGAGGTC CGGGGCCTCC TCGCCAAGGA CCTCGCCGTC	1080
TTGGCCTCGA GGGAGGGGCT AGACCTCGTG CCCGGGGACG ACCCATGCT CCTCGCCTAC	1140
CTCCTGGACC CCTCCAACAC CACCCCGAG GGGGTGGCGC GGCGCTACGG GGGGGAGTGG	1200
ACGGAGGACG CCGCCCACCG GGCCCTCCTC TCGGAGAGGC TCCATCGGAA CCTCCTTAAG	1260
CGCCTCGAGG GGGAGGAGAA GCTCCTTTGG CTCTACCACG AGGTGGAAAA GCCCCCTCTC	1320
CGGTCCTGG CCCACATGGA GGCCACCGGG GTACGGCTGG ACGTGGCCTA CCTTCAGGCC	1380
CTTCCCTGG AGCTTGCGGA GGAGATCCGC CGCCTCGAGG AGGAGGTCTT CCGCTTGGCG	1440
GGCCACCCCT TCAACCTCAA CTCCCGGGAC CAGCTGGAAA GGGTGCTCTT TGACGAGCTT	1500

AGGCTTCCCG CCTTGGGGAA GACGCAAAAG ACAGGCAAGC GCTCCACCAG CGCCGCGGTG 1560  
 CTGGAGGCCC TACGGGAGGC CCACCCCATC GTGGAGAAGA TCCTCCAGCA CCGGGAGCTC 1620  
 ACCAAGCTCA AGAACACCTA CGTGGACCCC CTCCCAAGCC TCGTCCACCC GAGGACGGGC 1680  
 CGCCTCCACA CCCGCTTCAA CCAGACGGCC ACGGCCACGG GGAGGCTTAG TAGCTCCGAC 1740  
 CCCAACCTGC AGAACATCCC CGTCCGCACC CCCTTGGGCC AGAGGATCCG CCGGGCCTTC 1800  
 GTGGCCGAGG CGGGTTGGGC GTTGGTGGCC CTGGACTATA GCCAGATAGA GCTCCGCGTC 1860  
 CTCGCCCACC TCTCCGGGGA CGAAAACCTG ATCAGGGTCT TCCAGGAGGG GAAGGACATC 1920  
 CACACCCAGA CCGCAAGCTG GATGTTCCGC GTCCCCCCGG AGGCCGTGGA CCCCTGATG 1980  
 CGCCGGGCGG CCAAGACGGT GAACTTCGGC GTCTCTACG GCATGTCCGC CCATAGGCTC 2040  
 TCCCAGGAGC TTGCCATCCC CTACGAGGAG GCGGTGGCCT TTATAGAGGC TACTTCCAAA 2100  
 GCTTCCCCAA GGTGCGGGCC TGGATAGAAA AGACCCTGGA GGAGGGGAGG AAGCGGGGCT 2160  
 ACGTGGAAC CCTCTTCGGA AGAAGGCGCT ACGTGCCCGA CCTCAACGCC CGGGTGAAGA 2220  
 GCGTCAGGA GGCCGCGGAG CGCATGGCCT TCAACATGCC CGTCCAGGGC ACCGCCGCCG 2280  
 ACCTCATGAA GCTCGCCATG GTGAAGCTCT TCCCCGCCT CCGGGAGATG GGGGCCCGCA 2340  
 TGCTCCTCCA GGTCCACGAC GAGCTCCTCC TGGAGGCCCC CCAAGCGCGG GCCGAGGAGG 2400  
 TGGCGGCTTT GGCCAAGGAG GCCATGGAGA AGGCCTATCC CCTCGCCGTG CCCCTGGAGG 2460  
 TGGAGGTGGG GATGGGGGAG GACTGGCTTT CCGCCAAGGG TTAG 2504

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 832 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu  
 1 5 10 15  
 Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly  
 20 25 30  
 Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala  
 35 40 45  
 Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val  
 50 55 60  
 Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly  
 65 70 75 80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu  
 85 90 95  
 Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu Glu  
 100 105 110  
 Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys  
 115 120 125  
 Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp  
 130 135 140  
 Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly  
 145 150 155 160  
 Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro  
 165 170 175  
 Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp Asn  
 180 185 190  
 Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu Leu  
 195 200 205  
 Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg Leu  
 210 215 220  
 Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu Lys  
 225 230 235 240  
 Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val  
 245 250 255  
 Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe  
 260 265 270  
 Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu  
 275 280 285  
 Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly  
 290 295 300  
 Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp  
 305 310 315 320  
 Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro  
 325 330 335  
 Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu  
 340 345 350  
 Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro  
 355 360 365  
 Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn  
 370 375 380  
 Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu  
 385 390 395 400  
 Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu  
 405 410 415

Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu  
420 425 430

Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly  
435 440 445

Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala  
450 455 460

Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His  
465 470 475 480

Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp  
485 490 495

Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg  
500 505 510

Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile  
515 520 525

Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr  
530 535 540

Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu  
545 550 555 560

His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser  
565 570 575

Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln  
580 585 590

Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala  
595 600 605

Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly  
610 615 620

Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr  
625 630 635 640

Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro  
645 650 655

Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly  
660 665 670

Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu  
675 680 685

Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg  
690 695 700

Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val  
705 710 715 720

Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg  
725 730 735

Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro  
740 745 750



Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu  
755 760 765  
Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His  
770 775 780  
Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala  
785 790 795 800  
Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro  
805 810 815  
Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu  
820 825 830

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 831 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val  
1 5 10 15  
Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu  
20 25 30  
Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys  
35 40 45  
Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Val Val Val Val Val  
50 55 60  
Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu Ala Tyr  
65 70 75 80  
Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu Ala  
85 90 95  
Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Val Arg Leu Glu Val  
100 105 110  
Pro Gly Phe Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Arg Ala  
115 120 125  
Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Asp Leu  
130 135 140  
Tyr Gln Leu Leu Ser Glu Arg Ile Ala Ile Leu His Pro Glu Gly Tyr  
145 150 155 160  
Leu Ile Thr Pro Ala Trp Leu Tyr Glu Lys Tyr Gly Leu Arg Pro Glu  
165 170 175  
Gln Trp Val Asp Tyr Arg Ala Leu Ala Gly Asp Pro Ser Asp Asn Ile  
180 185 190

Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Gln Arg Leu Ile Arg  
 195 200 205  
 Glu Trp Gly Ser Leu Glu Asn Leu Phe Gln His Leu Asp Gln Val Lys  
 210 215 220  
 Pro Ser Leu Arg Glu Lys Leu Gln Ala Gly Met Glu Ala Leu Ala Leu  
 225 230 235 240  
 Ser Arg Lys Leu Ser Gln Val His Thr Asp Leu Pro Leu Glu Val Asp  
 245 250 255  
 Phe Gly Arg Arg Arg Thr Pro Asn Leu Glu Gly Leu Arg Ala Phe Leu  
 260 265 270  
 Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu Glu  
 275 280 285  
 Gly Pro Lys Ala Ala Glu Glu Ala Pro Trp Pro Pro Glu Gly Ala  
 290 295 300  
 Phe Leu Gly Phe Ser Phe Ser Arg Pro Glu Pro Met Trp Ala Glu Leu  
 305 310 315 320  
 Leu Ala Leu Ala Gly Ala Trp Glu Gly Arg Leu His Arg Ala Gln Asp  
 325 330 335  
 Pro Leu Arg Gly Leu Arg Asp Leu Lys Gly Val Arg Gly Ile Leu Ala  
 340 345 350  
 Lys Asp Leu Ala Val Leu Ala Leu Arg Glu Gly Leu Asp Leu Phe Pro  
 355 360 365  
 Glu Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn Thr  
 370 375 380  
 Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu Asp  
 385 390 395 400  
 Ala Gly Glu Arg Ala Leu Leu Ala Glu Arg Leu Phe Gln Thr Leu Lys  
 405 410 415  
 Glu Arg Leu Lys Gly Glu Glu Arg Leu Leu Trp Leu Tyr Glu Glu Val  
 420 425 430  
 Glu Lys Pro Leu Ser Arg Val Leu Ala Arg Met Glu Ala Thr Gly Val  
 435 440 445  
 Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Val Glu Ala  
 450 455 460  
 Glu Val Arg Gln Leu Glu Glu Glu Val Phe Arg Leu Ala Gly His Pro  
 465 470 475 480  
 Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu  
 485 490 495  
 Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser  
 500 505 510  
 Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val  
 515 520 525



(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 834 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

Met Glu Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1      5      10      15
Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly
20      25      30
Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
35      40      45
Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe
50      55      60
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu
65      70      75      80
Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
85      90      95
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu
100     105     110
Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys
115     120     125
Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg
130     135     140
Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu
145     150     155     160
Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg
165     170     175
Pro Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp
180     185     190
Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Leu Lys Leu
195     200     205
Leu Lys Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg
210     215     220
Val Lys Pro Glu Asn Val Arg Glu Lys Ile Lys Ala His Leu Glu Asp
225     230     235     240
Leu Arg Leu Ser Leu Glu Leu Ser Arg Val Arg Thr Asp Leu Pro Leu
245     250     255
Glu Val Asp Leu Ala Gln Gly Arg Glu Pro Asp Arg Glu Gly Leu Arg
260     265     270

```

Ala Phe Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly  
275 280 285

Leu Leu Glu Ala Pro Ala Pro Leu Glu Glu Ala Pro Trp Pro Pro Pro  
290 295 300

Glu Gly Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp  
305 310 315 320

Ala Glu Leu Lys Ala Leu Ala Ala Cys Arg Asp Gly Arg Val His Arg  
325 330 335

Ala Ala Asp Pro Leu Ala Gly Leu Lys Asp Leu Lys Glu Val Arg Gly  
340 345 350

Leu Leu Ala Lys Asp Leu Ala Val Leu Ala Ser Arg Glu Gly Leu Asp  
355 360 365

Leu Val Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro  
370 375 380

Ser Asn Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp  
385 390 395 400

Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg  
405 410 415

Asn Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr  
420 425 430

His Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala  
435 440 445

Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu  
450 455 460

Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala  
465 470 475 480

Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu  
485 490 495

Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly  
500 505 510

Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His  
515 520 525

Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys  
530 535 540

Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly  
545 550 555 560

Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu  
565 570 575

Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu  
580 585 590

Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu  
595 600 605

(2) INFORMATION FOR SEQ ID NO:7:

SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2502 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

\_\_\_\_\_

NNGGCGGTGN	TCGTGGTCTT	TGACGCCAAG	GCCCCCTCCT	TCCGCCACGA	GGCCTACGAG	240
GCCTACAAGG	CGGGCCGGGC	CCCCACCCCG	GAGGACTTTC	CCCGGCAGCT	CGCCCTCATC	300
AAGGAGCTGG	TGGACCTCCT	GGGGCTTGCG	CGCCTCGAGG	TCCCCGGCTA	CGAGGCGGAC	360
GACGTNCTGG	CCACCCTGGC	CAAGAAGGCG	GAAAAGGAGG	GGTACGAGGT	GCGCATCCTC	420
ACCGCCGACC	GCGACCTCTA	CCAGCTCCTT	TCCGACCGCA	TCGCCGTCCT	CCACCCCGAG	480
GGGTACCTCA	TCACCCCGGC	GTGGCTTTGG	GAGAAGTACG	GCCTGAGGCC	GGAGCAGTGG	540
GTGGACTACC	GGGCCCTGGC	GGGGGACCCC	TCCGACAACC	TCCCCGGGGT	CAAGGGCATC	600
GGGGAGAAGA	CCGCCCNAG	GCTCCTCNAG	GAGTGGGGGA	GCCTGAAAAA	CCTCCTCAAG	660
AACCTGGACC	GGGTGAAGCC	CGCCNTCCGG	GAGAAGATCC	AGGCCACAT	GGANGACCTG	720
ANGCTCTCCT	GGGAGCTNTC	CCAGGTGCGC	ACCGACCTGC	CCCTGGAGGT	GGACTTCGCC	780
AAGNGGCGGG	AGCCCGACCG	GGAGGGGCTT	AGGGCCTTTC	TGGAGAGGCT	GGAGTTTGGC	840
AGCCTCTCCT	ACGAGTTCGG	CCTCCTGGAG	GGCCCCAAGG	CCCTGGAGGA	GGCCCCCTGG	900
CCCCCGCCGG	AAGGGGCTT	CGTGGGCTTT	GTCCTTTCCC	GCCCCGAGCC	CATGTGGGCC	960
GAGCTTCTGG	CCCTGGCCGC	CGCCAGGGAG	GGCCGGGTCC	ACCGGGCACC	AGACCCCTTT	1020
ANGGGCCTNA	GGGACCTNAA	GGAGGTGCGG	GGNCTCCTCG	CCAAGGACCT	GGCCGTTTTG	1080
GCCCTGAGGG	AGGGCCTNGA	CCTCNTGCCC	GGGGACGACC	CCATGCTCCT	CGCCTACCTC	1140
CTGGACCCCT	CCAACACCAC	CCCCGAGGGG	GTGGCCCCGG	GCTACGGGGG	GGAGTGGACG	1200
GAGGANGCGG	GGGAGCGGGC	CCTCCTNTCC	GAGAGGCTCT	TCCNGAACCT	NNNGCAGCGC	1260
CTTGAGGGGG	AGGAGAGGCT	CCTTTGGCTT	TACCAGGAGG	TGGAGAAGCC	CCTTTCCCGG	1320
GTCCTGGCCC	ACATGGAGGC	CACGGGGGTN	CGGCTGGACG	TGGCCTACCT	CCAGGCCCTN	1380
TCCCTGGAGG	TGGCGGAGGA	GATCCGCCGC	CTCGAGGAGG	AGGTCTTCCG	CCTGGCCGGC	1440
CACCCCTTCA	ACCTCAACTC	CCGGGACCAG	CTGGAAAGGG	TGCTCTTTGA	CGAGCTNGGG	1500
CTTCCCGCCA	TCGGCAAGAC	GGAGAAGACN	GGCAAGCGCT	CCACCAGCGC	CGCCGTGCTG	1560
GAGGCCCTNC	GNGAGGCCCA	CCCCATCGTG	GAGAAGATCC	TGCAGTACCG	GGAGCTCACC	1620
AAGCTCAAGA	ACACCTACAT	NGACCCCTTG	CCNGNCCTCG	TCCACCCAG	GACGGGCCGC	1680
CTCCACACCC	GCTTCAACCA	GACGGCCACG	GCCACGGGCA	GGCTTAGTAG	CTCCGACCCC	1740
AACCTGCAGA	ACATCCCCGT	CCGCACCCCN	CTGGGCCAGA	GGATCCGCCG	GGCCTTCGTG	1800
GCCGAGGAGG	GNTGGGTGTT	GGTGGCCCTG	GACTATAGCC	AGATAGAGCT	CCGGGTCCTG	1860
GCCCACCTCT	CCGGGGACGA	GAACCTGATC	CGGGTCTTCC	AGGAGGGGAG	GGACATCCAC	1920
ACCCAGACCG	CCAGCTGGAT	GTTGCGCGTC	CCCCCGGAGG	CCGTGGACCC	CCTGATGCGC	1980
CGGGCGGCCA	AGACCATCAA	CTTCGGGGTC	CTCTACGGCA	TGTCCGCCCA	CCGCCTCTCC	2040

CAGGAGCTTG CCATCCCCTA CGAGGAGGCG GTGGCCTTCA TTGAGCGCTA CTTCCAGAGC 2100  
 TTCCCCAAGG TGCGGGCCTG GATTGAGAAG ACCCTGGAGG AGGGCAGGAG GCGGGGGTAC 2160  
 GTGGAGACCC TCTTCGGCCG CCGGCGCTAC GTGCCCGACC TCAACGCCCC GGTGAAGAGC 2220  
 GTGCGGGAGG CGGCGGAGCG CATGGCCTTC AACATGCCCC TCCAGGGCAC CGCCGCCGAC 2280  
 CTCATGAAGC TGGCCATGGT GAAGCTCTTC CCCCGGCTNC AGGAAATGGG GGCCAGGATG 2340  
 CTCCTNCAGG TCCACGACGA GCTGGTCCTC GAGGCCCCCA AAGAGCGGGC GGAGGNGGTG 2400  
 GCCGCTTTGG CCAAGGAGGT CATGGAGGGG GTCTATCCCC TGGCCGTGCC CCTGGAGGTG 2460  
 GAGGTGGGGA TGGGGGAGGA CTGGCTCTCC GCCAAGGAGT AG 2502

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 833 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Xaa	Ala	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	1	5	10	15
Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	20	25	30	
Leu	Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	35	40	45	
Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Asp	Ala	Val	Xaa	Val	50	55	60	
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Glu	Ala	65	70	75	80
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu	85	90	95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Leu	Xaa	Arg	Leu	Glu	100	105	110	
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys	115	120	125	
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp	130	135	140	
Leu	Tyr	Gln	Leu	Leu	Ser	Asp	Arg	Ile	Ala	Val	Leu	His	Pro	Glu	Gly	145	150	155	160
Tyr	Leu	Ile	Thr	Pro	Ala	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	165	170	175	



Glu Gln Trp Val Asp Tyr Arg Ala Leu Xaa Gly Asp Pro Ser Asp Asn  
180 185 190

Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Xaa Lys Leu Leu  
195 200 205

Xaa Glu Trp Gly Ser Leu Glu Asn Leu Leu Lys Asn Leu Asp Arg Val  
210 215 220

Lys Pro Xaa Xaa Arg Glu Lys Ile Xaa Ala His Met Glu Asp Leu Xaa  
225 230 235 240

Leu Ser Xaa Xaa Leu Ser Xaa Val Arg Thr Asp Leu Pro Leu Glu Val  
245 250 255

Asp Phe Ala Xaa Arg Arg Glu Pro Asp Arg Glu Gly Leu Arg Ala Phe  
260 265 270

Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu  
275 280 285

Glu Xaa Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly  
290 295 300

Ala Phe Val Gly Phe Val Leu Ser Arg Pro Glu Pro Met Trp Ala Glu  
305 310 315 320

Leu Leu Ala Leu Ala Ala Ala Arg Xaa Gly Arg Val His Arg Ala Xaa  
325 330 335

Asp Pro Leu Xaa Gly Leu Arg Asp Leu Lys Glu Val Arg Gly Leu Leu  
340 345 350

Ala Lys Asp Leu Ala Val Leu Ala Leu Arg Glu Gly Leu Asp Leu Xaa  
355 360 365

Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn  
370 375 380

Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu  
385 390 395 400

Asp Ala Gly Glu Arg Ala Leu Leu Ser Glu Arg Leu Phe Xaa Asn Leu  
405 410 415

Xaa Xaa Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Xaa Glu  
420 425 430

Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala Thr Gly  
435 440 445

Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu Val Ala  
450 455 460

Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala Gly His  
465 470 475 480

Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp  
485 490 495

Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg  
500 505 510

Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile  
515 520 525

Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Asn Thr  
530 535 540

Tyr Ile Asp Pro Leu Pro Xaa Leu Val His Pro Arg Thr Gly Arg Leu  
545 550 555 560

His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser  
565 570 575

Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln  
580 585 590

Arg Ile Arg Arg Ala Phe Val Ala Glu Glu Gly Trp Xaa Leu Val Ala  
595 600 605

Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly  
610 615 620

Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr  
625 630 635 640

Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp Pro  
645 650 655

Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly  
660 665 670

Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu  
675 680 685

Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg  
690 695 700

Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val  
705 710 715 720

Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala Arg  
725 730 735

Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met Pro  
740 745 750

Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu  
755 760 765

Phe Pro Arg Leu Xaa Glu Met Gly Ala Arg Met Leu Leu Gln Val His  
770 775 780

Asp Glu Leu Val Leu Glu Ala Pro Lys Xaa Arg Ala Glu Xaa Val Ala  
785 790 795 800

Ala Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro  
805 810 815

Leu Glu Val Glu Val Gly Xaa Gly Glu Asp Trp Leu Ser Ala Lys Glu  
820 825 830

Xaa



CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCCTG 1560  
 GAGGCCCTCC GCGAGGCCCA CCCCATCGTG GAGAAGATCC TGCAGGCATG CAAGCTTGGC 1620  
 ACTGGCCGTC GTTTTACAAC GTCGTGA 1647

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2088 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATGAATTCGG GGATGCTGCC CCTCTTTGAG CCCAAGGGCC GGTCTCTCCT GGTGGACGGC 60  
 CACCACCTGG CCTACCGCAC CTTCCACGCC CTGAAGGGCC TCACCACCAG CCGGGGGGAG 120  
 CCGGTGCAGG CGGTCTACGG CTTGCGCAAG AGCCTCCTCA AGGCCCTCAA GGAGGACGGG 180  
 GACGCGGTGA TCGTGGTCTT TGACGCCAAG GCCCCCTCCT TCCGCCACGA GGCCTACGGG 240  
 GGGTACAAGG CGGGCCGGGC CCCACGCCG GAGGACTTTC CCCGGCAACT CGCCCTCATC 300  
 AAGGAGCTGG TGGACCTCCT GGGGCTGGCG CGCCTCGAGG TCCCGGGCTA CGAGGCGGAC 360  
 GACGTCTTGG CCAGCCTGGC CAAGAAGGCG GAAAAGGAGG GCTACGAGGT CCGCATCCTC 420  
 ACCGCCGACA AAGACCTTTA CCAGCTCCTT TCCGACCGCA TCCACGTCCT CCACCCCGAG 480  
 GGGTACCTCA TCACCCCGGC CTGGCTTTGG GAAAAGTACG GCCTGAGGCC CGACCAGTGG 540  
 GCCGACTACC GGGCCCTGAC CGGGGACGAG TCCGACAACC TTCCCGGGGT CAAGGGCATC 600  
 GGGGAGAAGA CGGCAGGAA GCTTCTGGAG GAGTGGGGGA GCCTGGAAGC CCTCCTCAAG 660  
 AACCTGGACC GGCTGAAGCC CGCCATCCGG GAGAAGATCC TGGCCACAT GGACGATCTG 720  
 AAGCTCTCCT GGGACCTGGC CAAGGTGCGC ACCGACCTGC CCCTGGAGGT GGACTTCGCC 780  
 AAAAGGCGGG AGCCCGACCG GGAGAGGCTT AGGGCCTTTC TGGAGAGGCT TGAGTTTGGC 840  
 AGCCTCCTCC ACGAGTTTCG CTTTCTGGAA AGCCCCAAGG CCCTGGAGGA GGCCCCCTGG 900  
 CCCCCGCCGG AAGGGGCCTT CGTGGGCTTT GTGCTTTCCC GCAAGGAGCC CATGTGGGCC 960  
 GATCTTCTGG CCCTGGCCGC CGCCAGGGGG GGCCGGGTCC ACCGGGCCCC CGAGCCTTAT 1020  
 AAAGCCCTCA GGGACCTGAA GGAGGCGCGG GGGCTTCTCG CCAAAGACCT GAGCGTTCTG 1080  
 GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GCGACGACC CCATGCTCCT CGCCTACCTC 1140  
 CTGGACCCTT CCAACACCAC CCCCAGGGGG GTGGCCCGGC GCTACGGCGG GGAGTGGACG 1200  
 GAGGAGGCGG GGGAGCGGGC CGCCCTTTCC GAGAGGCTCT TCGCCAACCT GTGGGGGAGG 1260  
 CTTGAGGGGG AGGAGAGGCT CTTTGGCTT TACCGGGAGG TGGAGAGGCC CCTTCCGCT 1320

GTCCTGGCCC	ACATGGAGGC	CACGGGGGTG	CGCCTGGACG	TGGCCTATCT	CAGGGCCTTG	1380
TCCCTGGAGG	TGGCCGGGGA	GATCGCCCGC	CTCGAGGCCG	AGGTCTTCCG	CCTGGCCGGC	1440
CACCCCTTCA	ACCTCAACTC	CCGGGACCAG	CTGGAAAGGG	TCCTCTTTGA	CGAGCTAGGG	1500
CTTCCCGCCA	TCGGCAAGAC	GGAGAAGACC	GGCAAGCGCT	CCACCAGCGC	CGCCGTCTTG	1560
GAGGCCCTCC	GCGAGGCCCA	CCCCATCGTG	GAGAAGATCC	TGCAGTACCG	GGAGCTCACC	1620
AAGCTGAAGA	GCACCTACAT	TGACCCCTTG	CCGGACCTCA	TCCACCCAG	GACGGGCCGC	1680
CTCCACACCC	GCTTCAACCA	GACGGCCACG	GCCACGGGCA	GGCTAAGTAG	CTCCGATCCC	1740
AACCTCCAGA	ACATCCCCGT	CCGCACCCCG	CTTGGGCAGA	GGATCCGCCG	GGCCTTCATC	1800
GCCGAGGAGG	GGTGGCTATT	GGTGGCCCTG	GACTATAGCC	AGATAGAGCT	CAGGGTGCTG	1860
GCCCACCTCT	CCGGCGACGA	GAACCTGATC	CGGGTCTTCC	AGGAGGGGCG	GGACATCCAC	1920
ACGGAGACCG	CCAGCTGGAT	GTTCGGCGTC	CCCCGGGAGG	CCGTGGACCC	CCTGATGCGC	1980
CGGGCGGCCA	AGACCATCAA	CTTCGGGGTC	CTCTACGGCA	TGTCGGCCCA	CCGCCTCTCC	2040
CAGGAGCTAG	CTAGCCATCC	CTTACGAGGA	GGCCCAGGCC	TTCAATTGA		2088

(i) SEQUENCE CHARACTERISTICS:

(ii) MOLECULE TYPE: DNA (genomic)

ATGAATTCGG	GGATGCTGCC	CCTCTTTGAG	CCCAAGGGCC	GGGTCCTCCT	GGTGGACGGC	60
CACCACCTGG	CCTACCGCAC	CTTCCACGCC	CTGAAGGGCC	TCACCACCAG	CCGGGGGGGAG	120
CCGGTGCAGG	CGGTCTACGG	CTTCGCCAAG	AGCCTCCTCA	AGGCCCTCAA	GGAGGACGGG	180
GACGCGGTGA	TCGTGGTCTT	TGACGCCAAG	GCCCCCTCCT	TCCGCCACGA	GGCCTACGGG	240
GGGTACAAGG	CGGGCCGGGC	CCCCACGCCG	GAGGACTTTC	CCCGGCAACT	CGCCCTCATC	300
AAGGAGCTGG	TGGACCTCCT	GGGGCTGGCG	CGCCTCGAGG	TCCCGGGCTA	CGAGGCGGAC	360
GACGTCCTGG	CCAGCCTGGC	CAAGAAGGCG	GAAAAGGAGG	GCTACGAGGT	CCGCATCCTC	420
ACCGCCGACA	AAGACCTTTA	CCAGCTTCTT	TCCGACCGCA	TCCACGTCCT	CCACCCCGAG	480
GGGTACCTCA	TCACCCCGGC	CTGGCTTTGG	GAAAAGTACG	GCCTGAGGCC	CGACCAGTGG	540
GCCGACTACC	GGGCCCTGAC	CGGGGACGAG	TCCGACAACC	TTCCCGGGGT	CAAGGGCATC	600
GGGGAGAAGA	CGGCGAGGAA	GCTTCTGGAG	GAGTGGGGGA	GCCTGGAAGC	CCTCCTCAAG	660
AACCTGGACC	GGCTGAAGCC	CGCCATCCGG	GAGAAGATCC	TGGCCACAT	GGACGATCTG	720

AAGCTCTCCT	GGGACCTGGC	CAAGGTGCGC	ACCGACCTGC	CCCTGGAGGT	GGACTTCGCC	780
AAAAGGCGGG	AGCCCGACCG	GGAGAGGCTT	AGGGCCTTTC	TGGAGAGGCT	TGAGTTTGGC	840
AGCCTCCTCC	ACGAGTTCGG	CCTTCTGGAA	AGCCCCAAGT	CATGGAGGGG	GTGTATCCCC	900
TGGCCGTGCC	CCTGGAGGTG	GAGGTGGGGA	TAGGGGAGGA	CTGGCTCTCC	GCCAAGGAGT	960
GA						962

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1600 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGAATTCG	GGGATGCTGC	CCCTCTTTGA	GCCCAAGGGC	CGGGTCCTCC	TGGTGGACGG	60
CCACCACCTG	GCCTACCGCA	CCTTCCACGC	CCTGAAGGGC	CTCACCACCA	GCCGGGGGGA	120
GCCGGTGCAG	GCGGTCTACG	GCTTCGCCAA	GAGCCTCCTC	AAGGCCCTCA	AGGAGGACGG	180
GGACGCGGTG	ATCGTGGTCT	TTGACGCCAA	GGCCCCCTCC	TTCCGCCACG	AGGCCTACGG	240
GGGGTACAAG	GCGGGCCGGG	CCCCACGCC	GGAGGACTTT	CCCCGGCAAC	TCGCCCTCAT	300
CAAGGAGCTG	GTGGACCTCC	TGGGGCTGGC	GCGCCTCGAG	GTCCCGGGCT	ACGAGGCGGA	360
CGACGTCCTG	GCCAGCCTGG	CCAAGAAGGC	GGAAAAGGAG	GGCTACGAGG	TCCGCATCCT	420
CACCGCCGAC	AAAGACCTTT	ACCAGCTCCT	TTCGACCGC	ATCCACGTCC	TCCACCCCGA	480
GGGGTACCTC	ATCACCCCGG	CCTGGCTTTG	GGAAAAGTAC	GGCCTGAGGC	CCGACCAGTG	540
GGCCGACTAC	CGGGCCCTGA	CCGGGGACGA	GTCCGACAAC	CTTCCCGGGG	TCAAGGGCAT	600
CGGGGAGAAG	ACGGCGAGGA	AGCTTCTGGA	GGAGTGGGGG	AGCCTGGAAG	CCCTCCTCAA	660
GAACCTGGAC	CGGCTGAAGC	CCGCCATCCG	GGAGAAGATC	CTGGCCCACA	TGGACGATCT	720
GAAGCTCTCC	TGGGACCTGG	CCAAGGTGCG	CACCGACCTG	CCCCTGGAGG	TGGACTTCGC	780
CAAAGGCGG	GAGCCCGACC	GGGAGAGGCT	TAGGGCCTTT	CTGGAGAGGC	TTGAGTTTGG	840
CAGCCTCCTC	CACGAGTTCG	GCCTTCTGGA	AAGCCCCAAG	ATCCGCCGGG	CCTTCATCGC	900
CGAGGAGGGG	TGGCTATTGG	TGGCCCTGGA	CTATAGCCAG	ATAGAGCTCA	GGGTGCTGGC	960
CCACCTCTCC	GGCGACGAGA	ACCTGATCCG	GGTCTTCCAG	GAGGGGCGGG	ACATCCACAC	1020
GGAGACCGCC	AGCTGGATGT	TCGGCGTCCC	CCGGGAGGCC	GTGGACCCCC	TGATGCGCCG	1080
GGCGGCCAAG	ACCATCAACT	TCGGGGTCCT	CTACGGCATG	TCGGCCCACC	GCCTCTCCCA	1140
GGAGCTAGCC	ATCCCTTACG	AGGAGGCCCA	GGCCTTCATT	GAGCGCTACT	TTCAGAGCTT	1200

CCCCAAGGTG CGGGCCTGGA TTGAGAAGAC CCTGGAGGAG GGCAGGAGGC GGGGGTACGT 1260  
 GGAGACCCTC TTCGGCCGCC GCCGCTACGT GCCAGACCTA GAGGCCCGGG TGAAGAGCGT 1320  
 GCGGGAGGCG GCCGAGCGCA TGGCCTTCAA CATGCCCGTC CGGGGCACCG CCGCCGACCT 1380  
 CATGAAGCTG GCTATGGTGA AGCTCTTCCC CAGGCTGGAG GAAATGGGGG CCAGGATGCT 1440  
 CCTTCAGGTC CACGACGAGC TGGTCCTCGA GGCCCCAAAA GAGAGGGCGG AGGCCGTGGC 1500  
 CCGGCTGGCC AAGGAGGTCA TGGAGGGGGT GTATCCCCTG GCCGTGCCCC TGGAGGTGGA 1560  
 GGTGGGGATA GGGGAGGACT GGCTCTCCGC CAAGGAGTGA 1600

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CACGAATTCG GGGATGCTGC CCCTCTTTGA GCCCAA 36

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 34 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGAGATCTA TCACTCCTTG GCGGAGAGCC AGTC 34

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 91 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TAATACGACT CACTATAGGG AGACCGGAAT TCGAGCTCGC CCGGGCGAGC TCGAATTCCG 60

TGTATTCTAT AGTGTACCT AAATCGAATT C 91

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TAATACGACT CACTATAGGG

20

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GAATTCGATT TAGGTGACAC TATAGAA

27

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTAATCATGG TCATAGCTGG TAGCTTGCTA C

31

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGATCCTCTA GAGTCGACCT GCAGGCATGC CTACCTTGGT AG

42



(2) INFORMATION FOR SEQ ID NO:20:

(ii) MOLECULE TYPE: DNA (genomic)

GGATCCTCTA GAGTCGACCT GCAGGCATGC

(2) INFORMATION FOR SEQ ID NO:21:

(ii) MOLECULE TYPE: DNA (genomic)

ATGAATTCGG	GGATGCTGCC	CCTCTTTGAG	CCCAAGGGCC	GGGTCCTCCT	GGTGGACGGC	60
CACCACCTGG	CCTACCGCAC	CTTCCACGCC	CTGAAGGGCC	TCACCACCAG	CCGGGGGGGAG	120
CCGGTGCAGG	CGGTCTACGG	CTTCGCCAAG	AGCCTCCTCA	AGGCCCTCAA	GGAGGACGGG	180
GACGCGGTGA	TCGTGGTCTT	TGACGCCAAG	GCCCCCTCCT	TCCGCCACGA	GGCCTACGGG	240
GGGTACAAGG	CGGGCCGGGC	CCCCACGCCG	GAGGACTTTC	CCCGGCAACT	CGCCCTCATC	300
AAGGAGCTGG	TGGACCTCCT	GGGGCTGGCG	CGCCTCGAGG	TCCCGGGCTA	CGAGGCGGAC	360
GACGTCCTGG	CCAGCCTGGC	CAAGAAGGCG	GAAAAGGAGG	GCTACGAGGT	CCGCATCCTC	420
ACCGCCGACA	AAGACCTTTA	CCAGCTCCTT	TCCGACCGCA	TCCACGTCTT	CCACCCCGAG	480
GGGTACCTCA	TCACCCCGGC	CTGGCTTTTG	GAAAAGTACG	GCCTGAGGCC	CGACCAGTGG	540
GCCGACTACC	GGGCCCTGAC	CGGGGACGAG	TCCGACAACC	TTCCCGGGGT	CAAGGGCATC	600
GGGGAGAAGA	CGGCGAGGAA	GCTTCTGGAG	GAGTGGGGGA	GCCTGGAAGC	CCTCCTCAAG	660
AACCTGGACC	GGCTGAAGCC	CGCCATCCGG	GAGAAGATCC	TGGCCACAT	GGACGATCTG	720
AAGTCTCCT	GGGACCTGGC	CAAGGTGCGC	ACCGACCTGC	CCCTGGAGGT	GGACTTCGCC	780
AAAAGGCGGG	AGCCCGACCG	GGAGAGGCTT	AGGGCCTTTC	TGGAGAGGCT	TGAGTTTGGC	840
AGCTTCCTCC	ACGAGTTCGG	CCTTCTGGAA	AGCCCCAAGG	CCCTGGAGGA	GGCCCCCTGG	900
CCCCCGCCGG	AAGGGGCCTT	CGTGGGCTTT	GTGCTTTCCC	GCAAGGAGCC	CATGTGGGCC	960
GATCTTCTGG	CCCTGGCCGC	CGCCAGGGGG	GGCCGGGTCC	ACCGGGCCCC	CGAGCCTTAT	1020
AAAGCCCTCA	GGGACCTGAA	GGAGGCGCGG	GGGCTTCTCG	CCAAAAGACCT	GAGCGTTCTG	1080

GCCCTGAGGG AAGGCCTTGG CCTCCCGCCC GCGACGACC CCATGCTCCT CGCCTACCTC	1140
CTGGACCCTT CCAACACCAC CCCCAGGGGG GTGGCCCGGC GCTACGGCGG GGAGTGGACG	1200
GAGGAGGCGG GGGAGCGGGC CGCCCTTTCC GAGAGGCTCT TCGCCAACCT GTGGGGGAGG	1260
CTTGAGGGGG AGGAGAGGCT CCTTTGGCTT TACCGGGAGG TGGAGAGGCC CCTTTCCGCT	1320
GTCCTGGCCC ACATGGAGGC CACGGGGGTG CGCCTGGACG TGGCCTATCT CAGGGCCTTG	1380
TCCCTGGAGG TGGCCGGGGA GATCGCCCGC CTCGAGGCCG AGGTCTTCCG CCTGGCCGGC	1440
CACCCCTTCA ACCTCAACTC CCGGACCAG CTGGAAAGGG TCCTCTTTGA CGAGCTAGGG	1500
CTTCCCGCCA TCGGCAAGAC GGAGAAGACC GGCAAGCGCT CCACCAGCGC CGCCGTCCTG	1560
GAGGCCCTCC GCGAGGCCCA CCCCATCGTG GAGAAGATCC TGCAGTACCG GGAGCTCACC	1620
AAGCTGAAGA GCACCTACAT TGACCCCTTG CCGGACCTCA TCCACCCAG GACGGGCCGC	1680
CTCCACACCC GCTTCAACCA GACGGCCACG GCCACGGGCA GGCTAAGTAG CTCCGATCCC	1740
AACCTCCAGA ACATCCCCGT CCGCACCCCG CTTGGGCAGA GGATCCGCCG GGCCTTCATC	1800
GCCGAGGAGG GGTGGCTATT GGTGGCCCTG GACTATAGCC AGATAGAGCT CAGGGTGCTG	1860
GCCCACCTCT CCGGCGACGA GAACCTGATC CGGGTCTTCC AGGAGGGGCG GGACATCCAC	1920
ACGGAGACCG CCAGCTGGAT GTTCGGCGTC CCCCGGGAGG CCGTGGACCC CCTGATGCGC	1980
CGGGCGGCCA AGACCATCAA CTTCGGGGTC CTCTACGGCA TGTCGGCCCA CCGCTCTCC	2040
CAGGAGCTAG CCATCCCTTA CGAGGAGGCC CAGGCCTTCA TTGAGCGCTA CTTTCAGAGC	2100
TTCCCCAAGG TGCGGGCCTG GATTGAGAAG ACCCTGGAGG AGGGCAGGAG GCGGGGGTAC	2160
GTGGAGACCC TCTTCGGCCG CCGCCGCTAC GTGCCAGACC TAGAGGCCCG GGTGAAGAGC	2220
GTGCGGGAGG CGGCCGAGCG CATGGCCTTC AACATGCCCG TCCGGGGCAC CGCCGCCGAC	2280
CTCATGAAGC TGGCTATGGT GAAGCTCTTC CCCAGGCTGG AGGAAATGGG GGCCAGGATG	2340
CTCCTTCAGG TCCACGACGA GCTGGTCCTC GAGGCCCCAA AAGAGAGGGC GGAGGCCGTG	2400
GCCCGGCTGG CCAAGGAGGT CATGGAGGGG GTGTATCCCC TGGCCGTGCC CCTGGAGGTG	2460
GAGGTGGGGA TAGGGGAGGA CTGGCTCTCC GCCAAGGAGT GA	2502

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GATTTAGGTG AACTATAG

19

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 72 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGGACGAACA AGCGAGACAG CGACACAGGT ACCACATGGT ACAAGAGGCA AGAGAGACGA	60
CACAGCAGAA AC	72

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 70 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GTTTCTGCTG TGTCGTCTCT CTTGCCTCTT GTACCATGTG GTACCTGTGT CGCTGTCTCG	60
CTTGTTTCGTC	70

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GACGAACAAG CGAGACAGCG	20
-----------------------	----

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTTTCTGCTG TGTCGTCTCT CTTG	24
----------------------------	----

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 46 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCTCTTGTAC CATGTGGTAC CTGTGTCGCT GTCTCGCTTG TTCGTC

46

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 50 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ACACAGGTAC CACATGGTAC AAGAGGCAAG AGAGACGACA CAGCAGAAAC

50

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Asn Ser  
1 5 10 15

(2) INFORMATION FOR SEO ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 969 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATGGCTAGCA TGACTGGTGG ACAGCAAATG GGTCGGATCA ATTCGGGGAT GCTGCCCCCTC 60

TTTGAGCCCA AGGGCCGGGT CCTCCTGGTG GACGGCCACC ACCTGGCCTA CCGCACCTTC 120

CACGCCCTGA AGGGCCTCAC CACCAGCCGG GGGGAGCCGG TGCAGGCGGT CTACGGCTTC 180

GCCAAGAGCC TCCTCAAGGC CCTCAAGGAG GACGGGGACG CGGTGATCGT GGTCTTTGAC 240

GCCAAGGCCC CTCCTTCCG CCACGAGGCC TACGGGGGGT ACAAGGCGGG CCGGGCCCCC 300  
 ACGCCGGAGG ACTTTCCCCG GCAACTCGCC CTCATCAAGG AGCTGGTGGA CCTCCTGGGG 360  
 CTGGCGCGCC TCGAGGTCCC GGGCTACGAG GCGGACGACG TCCTGGCCAG CCTGGCCAAG 420  
 AAGGCGGAAA AGGAGGGCTA CGAGGTCCGC ATCCTCACCG CCGACAAAGA CCTTTACCAG 480  
 CTTCTTTCCG ACCGCATCCA CGTCCTCCAC CCCGAGGGGT ACCTCATCAC CCCGGCCTGG 540  
 CTTTGGGAAA AGTACGGCCT GAGGCCCGAC CAGTGGGCCG ACTACCGGGC CCTGACCGGG 600  
 GACGAGTCCG ACAACCTTCC CGGGGTCAAG GGCATCGGGG AGAAGACGGC GAGGAAGCTT 660  
 CTGGAGGAGT GGGGGAGCCT GGAAGCCCTC CTCAAGAACC TGGACCGGCT GAAGCCCGCC 720  
 ATCCGGGAGA AGATCCTGGC CCACATGGAC GATCTGAAGC TCTCCTGGGA CCTGGCCAAG 780  
 GTGCGCACCG ACCTGCCCCT GGAGGTGGAC TTCGCCAAAA GGCGGGAGCC CGACCGGGAG 840  
 AGGCTTAGGG CCTTTCTGGA GAGGCTTGAG TTTGGCAGCC TCCTCCACGA GTTCGGCCTT 900  
 CTGGAAGCC CCAAGTCATG GAGGGGTGT ATCCCTGGC CGTGCCCTG GAGGTGGAGG 960  
 TGGGGATAG 969

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 948 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ATGGCTAGCA TGA CTGGTGG ACAGCAAATG GGT CGGATCA ATT CGGGGAT GCTGCCCCCTC 60  
 TTTGAGCCCA AGGGCCGGGT CCTCCTGGTG GACGCCACC ACCTGGCCTA CCGCACCTTC 120  
 CACGCCCTGA AGGGCCTCAC CACCAGCCGG GGGGAGCCGG TGCAGGCGGT CTACGGCTTC 180  
 GCCAAGAGCC TCCTCAAGGC CCTCAAGGAG GACGGGGACG CGGTGATCGT GGTCTTTGAC 240  
 GCCAAGGCCC CTCCTTCCG CCACGAGGCC TACGGGGGGT ACAAGGCGGG CCGGGCCCCC 300  
 ACGCCGGAGG ACTTTCCCCG GCAACTCGCC CTCATCAAGG AGCTGGTGGA CCTCCTGGGG 360  
 CTGGCGCGCC TCGAGGTCCC GGGCTACGAG GCGGACGACG TCCTGGCCAG CCTGGCCAAG 420  
 AAGGCGGAAA AGGAGGGCTA CGAGGTCCGC ATCCTCACCG CCGACAAAGA CCTTTACCAG 480  
 CTTCTTTCCG ACCGCATCCA CGTCCTCCAC CCCGAGGGGT ACCTCATCAC CCCGGCCTGG 540  
 CTTTGGGAAA AGTACGGCCT GAGGCCCGAC CAGTGGGCCG ACTACCGGGC CCTGACCGGG 600  
 GACGAGTCCG ACAACCTTCC CGGGGTCAAG GGCATCGGGG AGAAGACGGC GAGGAAGCTT 660  
 CTGGAGGAGT GGGGGAGCCT GGAAGCCCTC CTCAAGAACC TGGACCGGCT GAAGCCCGCC 720

ATCCGGGAGA AGATCCTGGC CCACATGGAC GATCTGAAGC TCTCCTGGGA CCTGGCCAAG	780
GTGCGCACCG ACCTGCCCCCT GGAGGTGGAC TTCGCCAAAA GGCGGGAGCC CGACCGGGAG	840
AGGCTTAGGG CCTTTCTGGA GAGGCTTGAG TTTGGCAGCC TCCTCCACGA GTTCGGCCTT	900
CTGGAAAGCC CCAAGGCCGC ACTCGAGCAC CACCACCACC ACCACTGA	948

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 206 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGCCAGGGTT TTCCAGTCA CGACGTTGTA AAACGACGGC CAGTGAATTG TAATACGACT	60
CACTATAGGG CGAATTCGAG CTCGGTACCC GGGGATCCTC TAGAGTCGAC CTGCAGGCAT	120
GCAAGCTTGA GTATTCTATA GTGTCACCTA AATAGCTTGG CGTAATCATG GTCATAGCTG	180
TTTCCTGTGT GAAATTGTTA TCCGCT	206

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TTCTGGGTTC TCTGCTCTCT GGTCGCTGTC TCGCTTGTC GTC	43
--	----

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCTGTCTCGC TTGTTTCGTC	19
-----------------------	----

(2) INFORMATION FOR SEQ ID NO:35:

20

(ii) MOLECULE TYPE: DNA (genomic)

GACGAACAAG CGAGACAGCG

(2) INFORMATION FOR SEQ ID NO:36:

24

(ii) MOLECULE TYPE: DNA (genomic)

TTCTGGGTTC TCTGCTCTCT GGTC

(2) INFORMATION FOR SEQ ID NO:37:

43

(ii) MOLECULE TYPE: DNA (genomic)

GACGAACAAG CGAGACAGCG ACCAGAGAGC AGAGAACCCA GAA

(2) INFORMATION FOR SEQ ID NO:38:

23

(ii) MOLECULE TYPE: DNA (genomic)

ACCAGAGAGC AGAGAACCCA GAA

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AACAGCTATG ACCATGATTA C

21

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 60 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GTTCTCTGCT CTCTGGTCGC TGTCTCGCTT GTGAAACAAG CGAGACAGCG TGGTCTCTCG

60

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGAGAGACCA CGCTG

15

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 52 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TC

52



(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AGAAAGGAAG GGAAGAAAGC GAAAGG

26

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GACGGGGAAA GCCGGCGAAC G

21

(2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAAAGCCGGC GAACGTGGCG

20

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GGCGAACGTG GCGAGAAAGG A

21



(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:  
GCCGGCGAAC GTGGCGAGAA AGGA 24

(2) INFORMATION FOR SEQ ID NO:51:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:  
GGTTTTTCTT TGAGGTTTAG 20

(2) INFORMATION FOR SEQ ID NO:52:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:  
GCGACACTCC ACCATAGAT 19

(2) INFORMATION FOR SEQ ID NO:53:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:  
CTGTCTTCAC GCAGAAAGC 19

(2) INFORMATION FOR SEQ ID NO:54:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:  
GCACGGTCTA CGAGACCTC 19

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TAATACGACT CACTATAGGG

20

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 337 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: not relevant
  - (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: RNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGAAAGCUU GCAUGCCUGC AGGUCGACUC UAGAGGAUCU ACUAGUCAUA UGGAUUCUGU	60
CUUCACGCAG AAAGCGUCUG GCCAUGGCGU UAGUAUGAGU GUCGUGCAGC CUCCAGGACC	120
CCCCCUCCCG GGAGAGGCAU AGUGGUCUGC GGAACCGGUG AGUACACCGG AAUUGCCAGG	180
ACGACCGGGU CCUUUCUUGG AUAAACCCGC UCAAUGCCUG GAGAUUUGGG CGUGCCCCCG	240
CAAGACUGCU AGCCGAGUAG UGUUGGGUCG CGAAAGGCCU UGUGGUACUG CCUGAUAGGG	300
UGCCUGCGAG UGCCCCGGGA GGUCUCGUAG ACCGUGC	337

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
  - (B) LOCATION: 18
  - (C) IDENTIFICATION METHOD: experimental
  - (D) OTHER INFORMATION: /evidence= EXPERIMENTAL
- /note= "The N at this position indicates the presence of a fluorescein dye on an abasic linker."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCGGTCGTCC TGGCAATNCC

20

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GTTTATCCAA GAAAGGACCC GGTCC

25

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CAGGGTGAAG GGAAGAAGAA AGCGAAAGGT

30

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CAGGGGGAAG GGAAGAAGAA AGCGAAAGGT

30

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
  - (B) LOCATION: 1..2
  - (C) IDENTIFICATION METHOD: experimental
  - (D) OTHER INFORMATION: /evidence= EXPERIMENTAL
- /mod\_base= OTHER  
 /note= "The T residues at positions 1 and 2 are amino modified T residues."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:  
TTCTTTTCAC CAGCGAGACG GG 22

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:  
ATTGGGCGCC AGGGTGGTTT TT 22

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 53 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:  
CCCGTCTCGC TGGTGAAAAG AAAAACCACC CTGGCGCCCA ATACGCAAAC CGC 53

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:  
GAATTCGATT TAGGTGACAC TATAGAATAC A 31

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:  
CCTTTGCTT TCTTCCCTTC CTTTCTCGCC ACGTTCGCC GC 42

## (2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GCCGGCGAAC GTGGCGAGAA AGGA

24

## (2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CAGAAGGAAG GGAAGAAAGC GAAAGG

26

## (2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

CAGGGGGAAG GGAAGAAAGC GAAAGG

26

## (2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CAGGGTACAG GGAAGAAAGC GAAAGG

26